

POISSON APPROXIMATION FOR TWO SCAN STATISTICS WITH RATES OF CONVERGENCE

XIAO FANG

DAVID SIEGMUND

*National University of Singapore and Stanford University,
Stanford University*

Abstract

As an application of Stein's method for Poisson approximation, we prove rates of convergence for the tail probabilities of two scan statistics that have been suggested for detecting local signals in sequences of independent random variables subject to possible change-points. Our formulation deals simultaneously with ordinary and with large deviations.

1 INTRODUCTION

Let $\{X_1, \dots, X_n\}$ be a sequence of random variables. A widely studied problem is to test the hypothesis that the X 's are independent and identically distributed against the alternative that for some $0 \leq i < j \leq n$, $\{X_{i+1}, \dots, X_j\}$ have a distribution that differs from the distribution of the other X 's. If $t := j - i$ is assumed known and the change in distribution is a shift in the mean, one common suggestion to detect the change is the statistic

$$M_{n,t} = \max_{1 \leq i \leq n-t+1} (X_i + \dots + X_{i+t-1}). \quad (1.1)$$

See Glaz, Naus and Wallenstein (2001) for an introduction to scan statistics.

When t is unknown but the distributions of the X 's are otherwise completely specified, the maximum log likelihood ratio statistic is

$$\max_{0 \leq i < j \leq n} (S_j - S_i) \quad (1.2)$$

where

$$S_i = \sum_{k=1}^i \log[f_1(X_k)/f_0(X_k)] \quad (1.3)$$

and f_0 (f_1 resp.) is the density function of X under the null hypothesis (alternative hypothesis resp.). Appropriate statistics when the distributions involve unknown parameters can be found, for example, in Yao (1993).

Asymptotic p values of test statistics (1.1) and (1.2) have been derived as $n \rightarrow \infty$ under certain distributional assumptions on X_1 . See, for example, Chan and Zhang (2007) and Siegmund (1988). The statistic (1.2) has also been studied for its role in queueing theory, where it has the interpretation of the maximum waiting time among the first n customers of a single server queue (cf. Iglehart (1971)). However, except for (1.1) in the special case when X_1 is a Bernoulli variable (cf. Arratia, Gordon and Waterman (1990)), and for (1.2) when the problem is scaled so that the probability is approximately zero (cf. Siegmund (1988)), the rate of convergence for these approximations is unknown. In this paper, we establish rate of convergence of tail approximations for both statistics (1.1) and (1.2) under the assumption that X_1 comes from an exponential family of distributions. The error in our approximation is relative error, hence is applicable when the probability is small as well as when it converges to a positive limit.

In practice simulations have been widely used to justify the accuracy of the approximations suggested here. The constants arising from our calculations are undoubtedly much too large to be an alternative source to justify use of the approximations in practice. We view the value of our approximations as providing understanding of the relations of various parameters involved in the approximations, and in particular the uniformity of the validity of the approximation for both large and ordinary deviations.

In the next section, we state our main results. Section 3 contains an introduction to our main technique, Stein's method, and the proof of our main results. We discuss related problems in Section 4.

2 MAIN RESULTS

2.1 Scan statistics with fixed window size

Let $\{X_1, \dots, X_n\}$ be independent, identically distributed random variables with distribution function F and $\mathbb{E}X_1 = \mu_0$. For a positive integer $t < n$, define

$$M_{n;t} = \max_{1 \leq i \leq n-t+1} (X_i + \dots + X_{i+t-1}).$$

For $a > \mu_0$ and $b := at$, we are interested in calculating approximately the probability $\mathbb{P}(M_{n;t} \geq b)$. The convergence rate for various suggested approximations for general X_1 is not known. In practice, Monte Carlo simulations have been widely used to justify the accuracy of theoretical results. In the following theorem, we provide a Poisson approximation with rate of convergence in the case that the distribution of X_1 can be imbedded in an exponential family of probability measures $\{F_\theta : \theta \in \Theta\}$ where

$$dF_\theta(x) = e^{\theta x - \Psi(\theta)} dF(x). \quad (2.1)$$

It is known that the mean and variance of F_θ are $\Psi'(\theta)$ and $\Psi''(\theta)$ respectively. We assume $F(x)$ is non-degenerate, i.e., $\Psi''(\theta) > 0$. In this paper, we

use $\mathbb{P}_\theta(\cdot)$ ($\mathbb{E}_\theta(\cdot)$ resp.) to denote the probability (expectation resp.) under which $X_1 \sim F_\theta$.

Theorem 2.1. *Let $\{X_1, \dots, X_n\}$ be independent, identically distributed random variables with distribution function F that can be imbedded in an exponential family, as above. Let $\mathbb{E}X_1 = \mu_0$. For integers $t < n$, define*

$$M_{n;t} = \max_{1 \leq i \leq n-t+1} (X_i + \dots + X_{i+t-1}).$$

Let $a > \mu_0$ be such that $\theta_a \in \Theta^\circ$, the interior of Θ , and let θ_a be defined by $\Psi'(\theta_a) = a$. Let $b = at$. Suppose that F is either arithmetic or for some $a' > a$ such that $\theta_{a'} \in \Theta^\circ$,

$$\sup_{\theta_a \leq \theta \leq \theta_{a'}} \int_{-\infty}^{\infty} |\varphi_\theta(t)|^\nu dt < \infty \text{ for some positive integer } \nu,$$

where φ_θ is the characteristic function of F_θ . Then for some constant C depending only on the exponential family (2.1), μ_0 , and a ,

$$|\mathbb{P}(M_{n;t} \geq b) - (1 - e^{-\lambda})| \leq C \left(\frac{(\log t)^2}{t} + \frac{(\log t \wedge \log(n-t))}{n-t} \right) (\lambda \wedge 1). \quad (2.2)$$

In the non-arithmetic case

$$\lambda = \frac{(n-t+1)e^{-(a\theta_a - \Psi(\theta_a))t}}{\theta_a \sigma_a (2\pi t)^{1/2}} \exp\left[-\sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}(e^{-\theta_a D_k^+})\right], \quad (2.3)$$

where $\sigma_a^2 = \Psi''(\theta_a)$, $D_k = \sum_{i=1}^k (X_i^a - X_i)$ and $\{X_i, X_i^a : i \geq 1\}$ are independent, $X_i \sim F$ and $X_i^a \sim F_{\theta_a}$. In the arithmetic case, we assume without loss of generality that X_i is integer-valued with span 1 where the span of an integer-valued random variable is defined to be the largest value of Δ such that

$$\sum_{k \in \mathbb{Z}} \mathbb{P}(X_i = k\Delta + w) = 1 \text{ for some } w \in \mathbb{Z}.$$

In this arithmetic case,

$$\lambda = \frac{(n-t+1)e^{-(a\theta_a - \Psi(\theta_a))t} e^{-\theta_a(\lceil b \rceil - b)}}{(1 - e^{-\theta_a})\sigma_a (2\pi t)^{1/2}} \exp\left(-\sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}(e^{-\theta_a D_k^+})\right)$$

where $\lceil b \rceil = \inf\{v \in \mathbb{Z} : v \geq b\}$.

Remark 2.1. The various expressions entering into λ will be explained below. Here it is important to note that provided $n-t$ and t are large the error of approximation is relative error, valid when n is relatively small, so $\lambda \sim 0$, and when λ is bounded away from 0. Although it is possible to trace through the proof of Theorem 2.1 and obtain a numerical value for the constant C in (2.2), it would be too large for practical purposes. Therefore, we do not pursue it here.

Remark 2.2. Arratia, Gordon and Waterman (1990) obtained a bound for $|\mathbb{P}(M_{n;t} \geq b) - (1 - e^{-\lambda})|$ for independent, identically distributed Bernoulli random variables. They do not restrict b to grow linearly in t with fixed slope. For fixed a , their bound is of form (cf. equations (11)–(13) of Arratia, Gordon and Waterman (1990))

$$C(e^{-ct} + \frac{t}{n})(\lambda \wedge 1).$$

Compared to their result, Theorem 2.1 applies to more general distributions and recovers typical limit theorems in the literature on scan statistics. As $t, n - t \rightarrow \infty$ Theorem 2.1 guarantees the relative error in (2.2) goes to 0. See, for example, Theorem 1 of Chan and Zhang (2007).

Remark 2.3. The infinite series appearing in the definition of λ is derived as an application of classical random walk results of Spitzer. It arises probabilistically in the proof of Theorem 2.1 in the form $\mathbb{E}[1 - \exp\{-\theta_a D_{\tau_+}\}]/\mathbb{E}(\tau_+)$, where $\tau_+ = \inf\{t : D_t > 0\}$. The series form is useful for numerical computation. For example, in the very special case that $X_1 \sim N(\mu_0, 1)$, we find that $X_1^a \sim N(a, 1)$ and in the definition of λ in (2.3),

$$\begin{aligned} & \exp\left(-\sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}(e^{-\theta_a D_k^+})\right) \\ &= \exp\left(-2 \sum_{k=1}^{\infty} \frac{1}{k} \Phi(-(a - \mu_0)\sqrt{k/2})\right) \\ &=: (a - \mu_0)^2 \nu(\sqrt{2}(a - \mu_0)) \end{aligned}$$

where the function $\nu(x)$ was defined in (4.38) of Siegmund (1985) and for small x satisfies $\exp(-cx) + o(x^2)$ for $c \approx 0.583$, while $\nu(x) \sim 2/x^2$ as $x \rightarrow \infty$. More generally, by Theorem 8.51 of Siegmund (1985), for the non-arithmetic case of Theorem 2.1,

$$\begin{aligned} & \exp\left(-\sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}(e^{-\theta_a D_k^+})\right) \\ &= (a - \mu_0) \theta_a \exp \left\{ -\frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{1}{\theta_a + i\lambda} - \frac{1}{i\lambda} \right) \right. \\ & \quad \left. \times \left[\log \left(\frac{1}{1 - g(\lambda)} \right) + \log(-i(a - \mu_0)\lambda) \right] d\lambda \right\}, \end{aligned}$$

where $g(\lambda) = \mathbb{E}e^{i\lambda D_1}$.

The case that X_i is integer-valued and a is the largest value X_i can take is not included by Theorem 2.1 because of the constraint $\theta_a \in \Theta^o$. The following corollary covers this case. The proof of it is simpler than the proof of Theorem 2.1 and the convergence rate we obtain is faster.

Corollary 2.2. *Let $\{X_1, \dots, X_n\}$ be independent, identically distributed random variables with distribution function F that can be imbedded in an exponential family, as in (2.1). Let $\mathbb{E}X_1 = \mu_0$. For integers $t < n$, define*

$$M_{n;t} = \max_{1 \leq i \leq n-t+1} (X_i + \dots + X_{i+t-1}).$$

Assume X_1 is integer-valued with span 1. Suppose $a = \sup\{x : p_x := \mathbb{P}(X_1 = x) > 0\}$ is finite. Let $b = at$. Then we have, with constants C and c depending only on p_a ,

$$|\mathbb{P}(M_{n;t} \geq b) - (1 - e^{-\lambda})| \leq C(\lambda \wedge 1)e^{-ct} \quad (2.4)$$

where

$$\lambda = (n - t)p_a^t(1 - p_a) + p_a^t.$$

Proof. Following the proof of Theorem 2.1, let

$$Y_1 = \mathbb{I}(X_1 = \dots = X_t = a)$$

and for $2 \leq \alpha \leq n - t + 1$,

$$Y_\alpha = \mathbb{I}(X_{\alpha-1} < a, X_\alpha = \dots = X_{\alpha+t-1} = a).$$

Then with $W = \sum_{\alpha=1}^{n-t+1} Y_\alpha$,

$$\mathbb{E}W = p_a^t + (n - t)p_a^t(1 - p_a) = \lambda.$$

Instead of (3.4), we have $\mathbb{P}(M_{n;t} \geq b) = \mathbb{P}(W \geq 1)$, and instead of (3.8), we have

$$|\mathbb{P}(W \geq 1) - (1 - e^{-\lambda})| \leq (1 \wedge \frac{1}{\lambda})(n - t + 1)(2t + 1)p_a^{2t}.$$

This proves the bound (2.4). □

2.2 Scan statistics with varying window size

Next we study the maximum log likelihood ratio statistic (1.2). Suppose in (1.3), $f_0(x) = dF_{\theta_0}(x)$ and $f_1(x) = dF_{\theta_1}(x)$ where $\{F_\theta : \theta \in \Theta\}$ is an exponential family as in (2.1) and $\theta_0 < \theta_1$. Then we have

$$S_i = \sum_{k=1}^i \log[f_1(X_k)/f_0(X_k)] = \sum_{k=1}^i (\theta_1 - \theta_0) \left(X_k - \frac{\Psi(\theta_1) - \Psi(\theta_0)}{\theta_1 - \theta_0} \right).$$

By appropriate change of parameters and a slight abuse of notation, studying (1.2) is equivalent to studying the following problem.

Let $\{X_1, \dots, X_n\}$ be independent, identically distributed random variables with distribution function F that can be imbedded in an exponential

family, as in (2.1). Let $\mathbb{E}X_1 = \mu_0 < 0$. Let $S_0 = 0$ and $S_i = \sum_{j=1}^i X_j$ for $1 \leq i \leq n$. Suppose there exist $\theta_1 > 0$ such that

$$\Psi'(\theta_1) = \mu_1, \quad \Psi(\theta_1) = 0. \quad (2.5)$$

For $b > 0$, we give an approximation to

$$p_{n,b} := \mathbb{P}\left(\max_{0 \leq i < j \leq n} (S_j - S_i) \geq b\right) \quad (2.6)$$

with an explicit error bound in the following theorem.

Theorem 2.3. *Under the above setting, assume $\theta_1 \in \Theta^o$ and F_{θ_1} is either arithmetic or satisfies*

$$\int_{-\infty}^{\infty} |\varphi_{\theta_1}(t)| dt < \infty \text{ where } \varphi_{\theta_1}(t) = \mathbb{E}_{\theta_1} e^{itX_1}.$$

Let $h(b) > 0$ be any function such that

$$h(b) \rightarrow \infty, \quad h(b) = O(b^{1/2}) \text{ as } b \rightarrow \infty.$$

Suppose $n - b/\mu_1 > b^{1/2}h(b)$. Then, for $p_{n,b}$ defined in (2.6), we have

$$|p_{n,b} - (1 - e^{-\lambda})| \leq C\lambda \left\{ \left(1 + \frac{b/h^2(b)}{n - b/\mu_1}\right) e^{-ch^2(b)} + \frac{b^{1/2}h(b)}{n - \frac{b}{\mu_1}} \right\} \quad (2.7)$$

where constants c, C only depending on the exponential family F_θ and θ_1 . In the non-arithmetic case,

$$\lambda = \left(n - \frac{b}{\mu_1}\right) \frac{e^{-\theta_1 b}}{\theta_1 \mu_1} \exp\left(-2 \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}_{\theta_1} e^{-\theta_1 S_k^+}\right).$$

In the arithmetic case, where we assume without loss of generality that X_i is integer-valued with span 1 and b is an integer,

$$\lambda = \left(n - \frac{b}{\mu_1}\right) \frac{e^{-\theta_1 b}}{(1 - e^{-\theta_1})\mu_1} \exp\left(-2 \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}_{\theta_1} e^{-\theta_1 S_k^+}\right).$$

Remark 2.4. We refer to Remark 2.3 for the numerical calculation of λ . Choosing $h(b) = b^{1/2}$, we get

$$|p_{n,b} - (1 - e^{-\lambda})| \leq C\lambda \left\{ e^{-cb} + \frac{b}{n} \right\}$$

from (2.7). By choosing $h(b) = C(\log b)^{1/2}$ with large enough C , we can see that the relative error in the Poisson approximation goes to zero under the conditions

$$b \rightarrow \infty, \quad (b \log b)^{1/2} \ll n - b/\mu_1 = O(e^{\theta_1 b}),$$

where $n - b/\mu_1 = O(e^{\theta_1 b})$ ensures that λ is bounded. For the smaller range (in which case $\lambda \rightarrow 0$)

$$b \rightarrow \infty, \quad \delta b \leq n - b/\mu_1 = o(e^{\frac{1}{2}\theta_1 b})$$

for some $\delta > 0$, Theorem 2 of Siegmund (1988) obtained more accurate estimates and the technique used is different from ours.

3 PROOFS

Before proving our main theorems, we first introduce our main tool: Stein's method. Stein's method was first introduced by Stein (1972) and further developed in Stein (1986) for normal approximation. Chen (1975) developed Stein's method for Poisson approximation, which has been widely applied especially in computational biology after the work by Arratia, Goldstein and Gordon (1990). We refer to Barbour and Chen (2005) for an introduction to Stein's method.

The following theorem provides a useful upper bound on the total variation distance between the distribution of a sum of locally dependent Bernoulli random variables and a Poisson distribution. The total variation distance between two distributions is defined as

$$d_{TV}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{A \subset \mathbb{R}} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|.$$

Theorem 3.1 (Arratia, Goldstein and Gordon (1990)). *Let $W = \sum_{\alpha \in A} Y_\alpha$ be a sum of Bernoulli random variables where A is the index set and $\mathbb{P}(Y_\alpha = 1) = 1 - \mathbb{P}(Y_\alpha = 0) = p_\alpha$. Let $\lambda = \sum_{\alpha \in A} p_\alpha$, and let $Poi(\lambda)$ denote the Poisson distribution with mean λ . Then,*

$$d_{TV}(\mathcal{L}(W), Poi(\lambda)) \leq (1 \wedge \frac{1}{\lambda})(b_1 + b_2 + b_3) \quad (3.1)$$

where given any neighborhood B_α for each α such that $\alpha \in B_\alpha \subset A$,

$$\begin{aligned} b_1 &:= \sum_{\alpha \in A} \sum_{\beta \in B_\alpha} p_\alpha p_\beta, \\ b_2 &:= \sum_{\alpha \in A} \sum_{\alpha \neq \beta \in B_\alpha} \mathbb{E}(Y_\alpha Y_\beta), \\ b_3 &:= \sum_{\alpha \in A} \mathbb{E} \left| \mathbb{E}(Y_\alpha - p_\alpha | \sigma(Y_\beta : \beta \notin B_\alpha)) \right|. \end{aligned} \quad (3.2)$$

Remark 3.1. If B_α is chosen such that X_α is independent of $\{X_\beta : \beta \notin B_\alpha\}$, then b_3 in (3.1) equals 0. Roughly speaking, in order for b_1 and b_2 to be small, the size of B_α has to be small and $\mathbb{E}(Y_\beta | Y_\alpha = 1) = o(1)$ for $\alpha \neq \beta \in B_\alpha$.

3.1 Proof of Theorem 2.1

In this proof, let C and c denote positive constants which may represent different values in different expressions. By choosing C to be large enough in (2.2), and using (cf. Theorem 1 and Theorem 6 of Petrov (1965))

$$\mathbb{P}(M_{n;t} \geq b) \leq (n-t+1)\mathbb{P}(X_1 + \dots + X_t \geq b) \sim (n-t+1)e^{-(a\theta_a - \Psi(\theta_a))t} / t^{1/2}, \quad (3.3)$$

where $x \sim y$ means x/y is bounded away from zero and infinity, the bound (2.2) holds true if t or $n-t$ is bounded. Therefore, in the sequel, we can assume t and $n-t$ to be larger than any given constant.

We embed the sequence $\{X_1, \dots, X_n\}$ into an infinite i.i.d. sequence $\{\dots, X_{-1}, X_0, X_1, \dots\}$. For each integer α , let

$$T_\alpha = X_\alpha + \dots + X_{\alpha+t-1}, \quad \tilde{Y}_\alpha = I(T_\alpha \geq b).$$

To avoid the clumping of 1's in the sequence (\tilde{Y}_α) which makes a Poisson approximation invalid, we define,

$$Y_\alpha = I(T_\alpha \geq b, T_{\alpha-1} < b, \dots, T_{\alpha-m} < b)$$

where $m \leq \sqrt{t}$ will be chosen later in (3.27). Let

$$W = \sum_{\alpha=1}^{n-t+1} Y_\alpha, \quad \lambda_1 = \mathbb{E}W = (n-t+1)\mathbb{E}Y_1.$$

In the following, we first bound $|\mathbb{P}(M_{n;t} \geq b) - \mathbb{P}(W \geq 1)|$, then bound the total variation distance between the distribution of W and $Poi(\lambda_1)$, finally we bound $|\lambda_1 - \lambda|$.

First, since $\{M_{n;t} \geq b\} \setminus \{W \geq 1\} \subset \cup_{\alpha=1}^m \{T_\alpha \geq b\}$, we have

$$0 \leq \mathbb{P}(M_{n;t} \geq b) - \mathbb{P}(W \geq 1) \leq m\mathbb{P}(X_1 + \dots + X_t \geq b). \quad (3.4)$$

Next, we apply Theorem 3.1 to bound the total variation distance between the distribution of W and $Poi(\lambda_1)$. For each $1 \leq \alpha \leq n-t+1$, define $B_\alpha = \{1 \leq \beta \leq n-t+1 : |\alpha - \beta| < t+m\}$. By definition of B_α , Y_α is independent of $\{Y_\beta : \beta \notin B_\alpha\}$. Therefore, b_3 in (3.2) equals zero. Since $|B_\alpha| < 2(t+m)$,

$$b_1 = \sum_{1 \leq \alpha \leq n-t+1} \sum_{\beta \in B_\alpha} \mathbb{E}Y_\alpha \mathbb{E}Y_\beta < 2(t+m)\lambda_1 \mathbb{E}Y_1.$$

By our definition of Y_α , for $1 \leq |\beta - \alpha| \leq m$, $\mathbb{E}Y_\alpha Y_\beta = 0$, and for $m < |\beta - \alpha| < t+m$, $\mathbb{E}Y_\alpha Y_\beta \leq \mathbb{E}\tilde{Y}_{\alpha \wedge \beta} \tilde{Y}_{\alpha \vee \beta}$. Therefore, by symmetry,

$$b_2 = \sum_{1 \leq \alpha \leq n-t+1} \sum_{\alpha \neq \beta \in B_\alpha} \mathbb{E}Y_\alpha Y_\beta < 2(n-t+1)\mathbb{E}\tilde{Y}_1 \sum_{\beta=m+2}^{m+t} \mathbb{P}(T_\beta \geq b | T_1 \geq b).$$

For $\beta \geq t + 1$,

$$\mathbb{P}(T_\beta \geq b | T_1 \geq b) = \mathbb{P}(T_1 \geq b).$$

Let a positive number $0 < \delta < 1 \wedge (a - \mu_0)/4$ be chosen such that

$$\Psi(\theta_a) - (\mu_0 + \delta)\theta_a > 0 \quad \text{and} \quad m < (a'' - a)t/\delta \quad (3.5)$$

where $a'' < a'$ will be chosen later. The first inequality above is possible because of the strict convexity of Ψ .

We observe that for $m + 2 \leq \beta \leq t$, $T_\beta \geq b$ and $X_{t+1} + \dots + X_{t+\beta-1} \leq (\mu_0 + \delta)(\beta - 1)$ together imply $X_\beta + \dots + X_t \geq at - (\mu_0 + \delta)(\beta - 1)$. Therefore,

$$\begin{aligned} & \sum_{\beta=m+2}^t \mathbb{P}(T_\beta \geq b | T_1 \geq b) \\ & \leq \sum_{\beta=m+2}^t \{ \mathbb{P}(X_{t+1} + \dots + X_{t+\beta-1} > (\mu_0 + \delta)(\beta - 1)) \\ & \quad + \mathbb{P}(X_\beta + \dots + X_t \geq at - (\mu_0 + \delta)(\beta - 1) | T_1 \geq b) \}. \end{aligned}$$

For the first term, we have

$$\begin{aligned} & \sum_{\beta=m+2}^t \mathbb{P}(X_{t+1} + \dots + X_{t+\beta-1} > (\mu_0 + \delta)(\beta - 1)) \\ & \leq \sum_{\beta=m+2}^t e^{-[\theta_{\mu_0+\delta}(\mu_0+\delta) - \Psi(\theta_{\mu_0+\delta})](\beta-1)} \\ & \leq \frac{e^{-[\theta_{\mu_0+\delta}(\mu_0+2\delta) - \Psi(\theta_{\mu_0+\delta})]m}}{1 - e^{-[\theta_{\mu_0+\delta}(\mu_0+2\delta) - \Psi(\theta_{\mu_0+\delta})]}}. \end{aligned} \quad (3.6)$$

By the bound on V on page 613 of Komlós and Tusnády (1975) and recalling that we have chosen δ such that $\Psi(\theta_a) - (\mu_0 + \delta)\theta_a > 0$,

$$\begin{aligned} & \sum_{\beta=m+2}^t \mathbb{P}(X_\beta + \dots + X_t \geq at - (\mu_0 + \delta)(\beta - 1) | T_1 \geq b) \\ & \leq C \sum_{\beta=m+2}^t e^{-[\Psi(\theta_a) - (\mu_0 + \delta)\theta_a](\beta-1)} \sqrt{\frac{t}{t - \beta + 1}} \\ & \leq C \frac{e^{-[\Psi(\theta_a) - (\mu_0 + \delta)\theta_a]m}}{(1 - e^{-[\Psi(\theta_a) - (\mu_0 + \delta)\theta_a]})}. \end{aligned} \quad (3.7)$$

Therefore,

$$b_2 \leq C(n - t + 1)\mathbb{P}(T_1 \geq b)[m\mathbb{P}(T_1 \geq b) + e^{-cm}].$$

By Theorem 3.1,

$$\begin{aligned} & |\mathbb{P}(W \geq 1) - (1 - e^{-\lambda_1})| \\ & \leq C(1 \wedge \frac{1}{\lambda_1})(n - t + 1)\mathbb{P}(T_1 \geq b)[t\mathbb{P}(T_1 \geq b) + e^{-cm}]. \end{aligned} \quad (3.8)$$

Finally, we calculate approximately $\mathbb{E}Y_1$. By symmetry, we can write

$$\begin{aligned} \mathbb{E}Y_1 &= \mathbb{I}(T_1 \geq b, T_2 < b, \dots, T_{m+1} < b) \\ &= \mathbb{E}\tilde{Y}_1(1 - \tilde{Y}_2) \dots (1 - \tilde{Y}_{m+1}) \\ &\leq \int_b^{b+m\delta} \mathbb{E}[(1 - \tilde{Y}_2) \dots (1 - \tilde{Y}_{m+1}) | S_t = s] d\mathbb{P}(S_t \leq s) + \mathbb{P}(S_t > b + m\delta) \end{aligned} \quad (3.9)$$

where $S_t = X_1 + \dots + X_t$. Observe that $T_1 = s$ and $T_{i+1} < b$ imply $T_1 - T_{i+1} = S_i - (S_{i+t} - S_t) > s - b$. Therefore, given $T_1 = s$, $(1 - \tilde{Y}_2) \dots (1 - \tilde{Y}_{m+1})$ is the indicator of the event that $\{\tilde{S}_i^{s/t} - S_i > s - b, 1 \leq i \leq m\}$ where $\tilde{S}_i^{s/t}$ is independent of S_i ,

$$\tilde{S}_i^{s/t} = \sum_{j=1}^i \tilde{X}_j^{s/t} \quad \text{and} \quad \mathcal{L}(\tilde{X}_i^{s/t} : 1 \leq i \leq m) = \mathcal{L}(X_i : 1 \leq i \leq m | S_t = s).$$

Note that the assumption $m < (a'' - a)t/\delta$ in (3.5) implies $a \leq s/t \leq a''$. It is known that when m is small compared to t , the conditional sequence $\{\tilde{X}_i^{s/t} : 1 \leq i \leq m\}$ behaves like an i.i.d. sequence $\{X_i^{s/t} : 1 \leq i \leq m\}$ where $X_i^{s/t}$ comes from the same exponential family (2.1) as X_i , but with a different parameter $\theta_{s/t}$. From the proof of Theorem 1.6 of Diaconis and Freedman (1988) and the assumption on the exponential family in the statement of the theorem, we have

$$d_{TV}(\mathcal{L}(\tilde{X}_i^{s/t} : 1 \leq i \leq m), \mathcal{L}(X_i^{s/t} : 1 \leq i \leq m)) \leq C \frac{m}{t}. \quad (3.10)$$

In fact, for the non-arithmetic case, since only the range of parameters $[a, a']$ enters into considerations, we do not need Condition 1.1 of Diaconis and Freedman (1988). In the following we verify their Conditions 1.2–1.4. By our assumptions for the non-arithmetic case, their Conditions 1.2 and 1.4 are satisfied for the range of parameters $[a, a']$. By $\int_{-\infty}^{\infty} |\varphi_{\theta_a}(t)|^v dt < \infty$, we have for $t \neq 0$, $|\varphi_{\theta_a}(t)| < 1$ and $|\varphi_{\theta_a}(t)| \rightarrow 0$ as $|t| \rightarrow \infty$. Therefore, there exists $M > 0$ such that $|\varphi_{\theta_a}(t)| < 1/2$ for $|t| > M$. This, together with the fact that

$$|\varphi_{\theta_{a+h}}(t) - \varphi_{\theta_a}(t)| \rightarrow 0 \text{ as } h \rightarrow 0^+ \text{ uniformly in } t$$

by the dominated convergence theorem, implies that there exists $a' > a'' > a$ such that

$$\sup_{\theta_a \leq \theta \leq \theta_{a''}} \sup_{|t| > \delta} |\varphi_{\theta}(t)| < 1 \text{ for all } \delta > 0.$$

Therefore, Conditions 1.2–1.4 of Diaconis and Freedman (1988) are satisfied for the range of parameters $[a, a'']$, which yields (3.10). The arithmetic case can be proved similarly.

By the likelihood ratio identity, for $b < s \leq b + m\delta$ and $m \geq \nu$,

$$\begin{aligned} & d_{TV}\left(\mathcal{L}(X_i^{s/t} : 1 \leq i \leq m), \mathcal{L}(X_i^a : 1 \leq i \leq m)\right) \\ & \leq \mathbb{E}_{\theta_{s/t}} I(S_m > t/m) + \mathbb{E}_{\theta_a} I(S_m > t/m) \\ & \quad + \mathbb{E}_{\theta_a} |e^{(\theta_{s/t} - \theta_a)S_m - m(\Psi(\theta_{s/t}) - \Psi(\theta_a))} - 1| I(S_m \leq t/m). \end{aligned}$$

For $s/t \in [a, a'']$, we have

$$\begin{aligned} |\theta_{s/t} - \theta_a| & \leq \sup_{\theta_a \leq \theta \leq \theta_{a''}} \frac{1}{\Psi''(\theta)} (s/t - a), \\ |\Psi(\theta_{s/t}) - \Psi(\theta_a)| & \leq \sup_{\theta_a \leq \theta \leq \theta_{a''}} |\Psi'(\theta)| (s/t - a). \end{aligned} \tag{3.11}$$

This implies that if $a < s/t \leq a + m\delta/t$, $S_m \leq t/m$ and $m \leq \sqrt{t}$, then

$$(\theta_{s/t} - \theta_a)S_m - m(\Psi(\theta_{s/t}) - \Psi(\theta_a)) \leq C.$$

Therefore, by Markov's inequality and the fact that $|e^t - 1| \leq Ct$ if t is bounded, for $b \leq s \leq b + m\delta$,

$$\begin{aligned} & d_{TV}\left(\mathcal{L}(X_i^{s/t} : 1 \leq i \leq m), \mathcal{L}(X_i^a : 1 \leq i \leq m)\right) \\ & \leq \frac{m}{t} (\mathbb{E}_{\theta_{s/t}} |S_m| + \mathbb{E}_{\theta_a} |S_m|) + C \mathbb{E}_{\theta_a} |(\theta_{s/t} - \theta_a)S_m - m(\Psi(\theta_{s/t}) - \Psi(\theta_a))| \\ & \leq Cm^2/t \end{aligned} \tag{3.12}$$

where in the last inequality we used $s/t \in [a, a'']$, $\mathbb{E}_{\theta} |S_m| \leq Cm$ for $\theta \in [\theta_a, \theta_{a''}]$ and (3.11). Therefore, with $D_i := S_i^a - S_i$ where $S_i^a = \sum_{j=1}^i X_j^a$ and X_j^a is defined in the statement of Theorem 2.1, we have by (3.9), (3.10) and (3.12),

$$\begin{aligned} \mathbb{E}Y_1 & \leq \int_b^{b+m\delta} [\mathbb{P}(D_i > s - b, 1 \leq i \leq m) + C \frac{m^2}{t}] d\mathbb{P}(S_t \leq s) \\ & \quad + \mathbb{P}(S_t > b + m\delta). \end{aligned} \tag{3.13}$$

Recalling $0 < \delta < 1 \wedge (a - \mu_0)/4$ above (3.5) so that $b \leq s \leq b + m\delta$ implies

$$s - b - m(a - \mu_0)/2 < m(\mu_0 - a)/4,$$

we have for $m_1 > m$,

$$\begin{aligned}
& \mathbb{P}(D_i > s - b, m_1 \geq i \geq 1) \\
&= \mathbb{P}(D_i > s - b, 1 \leq i \leq m) \mathbb{P}(D_i > s - b, m_1 \geq i > m | D_i > s - b, 1 \leq i \leq m) \\
&\geq \mathbb{P}(D_i > s - b, 1 \leq i \leq m) \mathbb{P}(D_i > s - b, m_1 \geq i > m) \\
&\geq \mathbb{P}(D_i > s - b, 1 \leq i \leq m) \mathbb{P}(D_i > s - b, i > m, D_m \geq m(a - \mu_0)/2) \\
&\geq \mathbb{P}(D_i > s - b, 1 \leq i \leq m) \\
&\quad \times \left\{ 1 - \mathbb{P}(D_m < m(a - \mu_0)/2) - \sum_{i=1}^{\infty} \mathbb{P}(D_i < m(\mu_0 - a)/4) \right\}.
\end{aligned} \tag{3.14}$$

The first inequality in (3.14) follows from the FKG inequality (cf. (1.7) of Karlin and Rinott (1980)) and the fact that $\mathbb{I}(D_i > s - b, 1 \leq i \leq m)$ and $\mathbb{I}(D_i > s - b, m_1 \geq i > m)$ are both increasing functions of $\{X_1^a - X_1, \dots, X_{m_1}^a - X_{m_1}\}$. Letting $m_1 \rightarrow \infty$, we have

$$\begin{aligned}
& \mathbb{P}(D_i > s - b, i \geq 1) \\
&\geq \mathbb{P}(D_i > s - b, 1 \leq i \leq m) \\
&\quad \times \left\{ 1 - \mathbb{P}(D_m < m(a - \mu_0)/2) - \sum_{i=1}^{\infty} \mathbb{P}(D_i < m(\mu_0 - a)/4) \right\}.
\end{aligned} \tag{3.15}$$

For $0 < r \leq \theta_a$,

$$\mathbb{E} \exp(-r D_i) = \exp \left\{ -i [\Psi(\theta_a) - \Psi(r) - \Psi(\theta_a - r)] \right\}.$$

By Taylor's expansion,

$$\Psi(\theta_a) - \Psi(\theta_a - r) = r a - \frac{r^2}{2} \Psi''(\theta_a - r_1), \quad -\Psi(r) = -r \mu_0 - \frac{r^2}{2} \Psi''(r_2)$$

where $0 \leq r_1, r_2 \leq r$. Therefore,

$$\begin{aligned}
& \mathbb{P}(D_m < m(a - \mu_0)/2) \\
&\leq \exp \left\{ -m \left[\Psi(\theta_a) - \Psi(r) - \Psi(\theta_a - r) - \frac{(a - \mu_0)r}{2} \right] \right\} \\
&= \exp \left\{ -m \left[\frac{r}{2}(a - \mu_0) - \frac{r^2}{2} (\Psi''(\theta_a - r_1) + \Psi''(r_2)) \right] \right\}.
\end{aligned}$$

Let

$$c_1 = \frac{a - \mu_0}{4\theta_a} \vee \max_{\theta \in \Theta: 0 \leq \theta \leq \theta_a} \Psi''(\theta).$$

Choosing $r = (a - \mu_0)/(4c_1)$, we have

$$\mathbb{P}(D_m < m(a - \mu_0)/2) \leq \exp \left\{ -\frac{(a - \mu_0)^2}{16c_1} m \right\}. \tag{3.16}$$

Similarly,

$$\mathbb{P}(D_i < m(\mu_0 - a)/4) \leq \exp \left\{ -\frac{(a - \mu_0)^2}{16c_1}i - \frac{(a - \mu_0)^2}{16c_1}m \right\}. \quad (3.17)$$

Applying (3.16) and (3.17) in (3.15), we obtain

$$\begin{aligned} & \mathbb{P}(D_i > s - b, 1 \leq i \leq m) \\ & \leq \mathbb{P}(D_i > s - b, i \geq 1) + 2 \exp \left\{ -\frac{(a - \mu_0)^2}{16c_1}m \right\} / \left(1 - \exp \left\{ -\frac{(a - \mu_0)^2}{16c_1} \right\} \right). \end{aligned}$$

Therefore, by (3.13),

$$\begin{aligned} & \mathbb{E}Y_\alpha - \lambda_2/(n - t + 1) \\ & \leq \left\{ \frac{2 \exp \left\{ -m(a - \mu_0)^2/(16c_1) \right\}}{1 - \exp \left\{ -(a - \mu_0)^2/(16c_1) \right\}} + C \frac{m^2}{t} + \frac{\mathbb{P}(S_t > b + m\delta)}{\mathbb{P}(S_t \geq b)} \right\} \mathbb{P}(S_t \geq b) \end{aligned} \quad (3.18)$$

where

$$\lambda_2 = (n - t + 1) \int_b^\infty \mathbb{P}(D_i > s - b, i \geq 1) d\mathbb{P}(S_t \leq s).$$

From the corollary on page 611 of Komlós and Tusnády (1975), and recalling that in proving (2.2), we can only consider those t larger than any given constant, we have

$$\mathbb{P}(T_1 > b + m\delta | T_1 \geq b) \leq C e^{-\theta_a m\delta}. \quad (3.19)$$

After proving a similar and easier lower bound of $\mathbb{E}Y_1$, we obtain, along with (3.19),

$$|\mathbb{E}Y_1 - \lambda_2/(n - t + 1)| \leq C \left[\frac{m^2}{t} + e^{-cm} \right] \mathbb{P}(S_t \geq b). \quad (3.20)$$

To calculate λ_2 , we first consider the non-arithmetic case of Theorem 2.1. By the proof of Theorem 2.7 of Woodroffe (1982), we have for $x \geq 0$,

$$\mathbb{P}(D_i > x, i \geq 1) = \frac{\mathbb{P}(D_{\tau_+} > x)}{\mathbb{E}\tau_+} \quad (3.21)$$

where $\tau_+ = \inf\{i \geq 1, D_i > 0\}$. Let $x_0 = \log t/\theta_a$. By change of variable and the likelihood ratio identity,

$$\begin{aligned} \lambda_2 &= (n - t + 1) \int_0^\infty \mathbb{P}(D_i > x, i \geq 1) d\mathbb{P}(S_t \leq b + x) \\ &= (n - t + 1) e^{-(a\theta_a - \Psi(\theta_a))t} \\ &\quad \times \int_0^{x_0} \mathbb{P}(D_i > x, i \geq 1) e^{-\theta_a x} d\mathbb{P}_{\theta_a}(S_t \leq b + x) \\ &\quad + O((n - t + 1) \mathbb{P}(S_t > b + x_0)). \end{aligned} \quad (3.22)$$

By the local central limit theorem (cf. Feller (1971)), uniformly for $0 \leq x \leq x_0$,

$$d\mathbb{P}_{\theta_a}(S_t \leq b+x) = \frac{1}{\sigma_a(2\pi t)^{1/2}} + O\left(\frac{(\log t)^2}{t^{3/2}}\right). \quad (3.23)$$

By (3.19) and (3.3),

$$\begin{aligned} \mathbb{P}(S_t > b+x_0) &= \mathbb{P}(S_t > b+x_0 | S_t \geq b) \mathbb{P}(S_t \geq b) \\ &\leq C e^{-\theta_a x_0} \frac{e^{-(a\theta_a - \Psi(\theta_a))t}}{\sqrt{t}} \leq C \frac{e^{-(a\theta_a - \Psi(\theta_a))t}}{\sqrt{t}} \frac{1}{t}. \end{aligned} \quad (3.24)$$

Applying (3.21), (3.23) and (3.24) in (3.22), we obtain

$$\begin{aligned} \lambda_2 &= \frac{(n-t+1)e^{-(a\theta_a - \Psi(\theta_a))t}}{(\mathbb{E}\tau_+)\sigma_a(2\pi t)^{1/2}} \\ &\quad \times \int_0^{x_0} \mathbb{P}(D_{\tau_+} > x) e^{-\theta_a x} \left(1 + O\left(\frac{(\log t)^2}{t}\right)\right) dx \\ &\quad + O((n-t+1)\mathbb{P}(S_t > b+x_0)) \\ &= \frac{(n-t+1)e^{-(a\theta_a - \Psi(\theta_a))t}}{(\mathbb{E}\tau_+)\sigma_a(2\pi t)^{1/2}} \\ &\quad \times \left[\int_0^\infty \mathbb{P}(D_{\tau_+} > x) e^{-\theta_a x} \left(1 + O\left(\frac{(\log t)^2}{t}\right)\right) dx + O\left(\frac{1}{t}\right) \right]. \end{aligned}$$

By the integration by parts formula,

$$\begin{aligned} &\frac{1}{\mathbb{E}\tau_+} \int_0^\infty \mathbb{P}(D_{\tau_+} > x) e^{-\theta_a x} dx \\ &= \frac{1}{\theta_a \mathbb{E}\tau_+} \left[1 - \mathbb{E}e^{-\theta_a D_{\tau_+}} \right] \\ &= \frac{1}{\theta_a \mathbb{E}\tau_+} \exp \left[- \sum_{k=1}^\infty k^{-1} \mathbb{E}(e^{-\theta_a D_k}, D_k > 0) \right] \\ &= \frac{1}{\theta_a} \exp \left[- \sum_{k=1}^\infty k^{-1} \mathbb{E}(e^{-\theta_a D_k^+}) \right] \end{aligned} \quad (3.25)$$

where we used the first equality in the proof of Corollary 2.7 of Woodroffe (1982) and Corollary 2.4 of Woodroffe (1982). Therefore,

$$\begin{aligned} \lambda_2 &= \frac{(n-t+1)e^{-(a\theta_a - \Psi(\theta_a))t}}{\theta_a \sigma_a(2\pi t)^{1/2}} \\ &\quad \times \exp \left[- \sum_{k=1}^\infty k^{-1} \mathbb{E}(e^{-\theta_a D_k^+}) \right] \left(1 + O\left(\frac{(\log t)^2}{t}\right)\right). \end{aligned} \quad (3.26)$$

Let

$$m = \lfloor C(\log t \wedge \log(n-t)) \rfloor \quad (3.27)$$

such that $e^{-cm} = O(\frac{1}{t} \vee \frac{1}{n-t})$ for the constants c in (3.8) and (3.20). Recall that in proving (2.2), we can only consider those t larger than any given number, thus (3.5) is satisfied. From (3.3),

$$\lambda \sim (n-t+1)\mathbb{P}(X_1 + \dots + X_t \geq b).$$

By (3.20) and (3.26),

$$|\lambda_1 - \lambda| \leq C\lambda \left[\frac{(\log t)^2}{t} + \frac{1}{n-t} \right].$$

By (3.4) and (3.8),

$$|\mathbb{P}(M_{n;t} \geq b) - (1 - e^{-\lambda_1})| \leq C(\lambda \wedge 1) \left[e^{-ct} + \frac{\log t \wedge \log(n-t)}{n-t} \right].$$

The bound (2.2) is proved by using the above two bounds for the cases $\lambda = O(1)$ and $\lambda \gg 1$ separately and using $|e^{-\lambda} - e^{-\lambda_1}| \leq |\lambda - \lambda_1|e^{-(\lambda \wedge \lambda_1)}$.

Next we consider the arithmetic case of Theorem 2.1. Without loss of generality, we assume X_1 is integer valued with span 1. The calculation of λ_2 is similar to the non-arithmetic case except that we have, for integers $0 \leq k \leq x_0$,

$$\mathbb{P}_{\theta_a}(S_t = \lceil b \rceil + k) = \frac{1}{\sigma_a(2\pi t)^{1/2}} + O\left(\frac{(\log t)^2}{t^{3/2}}\right)$$

and

$$\begin{aligned} & \sum_{k=0}^{\infty} \mathbb{P}(D_{\tau_+} > \lceil b \rceil - b + k) e^{-\theta_a(\lceil b \rceil - b + k)} \\ &= e^{-\theta_a(\lceil b \rceil - b)} \sum_{k=0}^{\infty} \mathbb{P}(D_{\tau_+} > k) e^{-\theta_a k} \\ &= \frac{e^{-\theta_a(\lceil b \rceil - b)}}{1 - e^{-\theta_a}} [1 - \mathbb{E}e^{-\theta_a D_{\tau_+}}]. \end{aligned} \tag{3.28}$$

Therefore, for the arithmetic case,

$$\begin{aligned} \lambda_2 &= \frac{(n-t+1)e^{-(a\theta_a - \Psi(\theta_a))t} e^{-\theta_a(\lceil b \rceil - b)}}{(1 - e^{-\theta_a})\sigma_a(2\pi t)^{1/2}} \\ &\quad \times \exp \left[- \sum_{n=1}^{\infty} n^{-1} \mathbb{E}(e^{-\theta_a D_n^+}) \right] \left(1 + O\left(\frac{(\log t)^2}{t}\right) \right). \end{aligned}$$

3.2 Proof of Theorem 2.3

Recall $S_i = \sum_{k=1}^i X_k$. Define $\tau_+ = \inf\{n \geq 1 : S_n > 0\}$ and

$$\tau_b := \inf\{n \geq 1 : S_n \geq b\}, \quad T_b := \inf\{n \geq 1 : S_n \notin [0, b)\}.$$

In this proof, let C and c denote positive constants which may represent different values in different expressions. If b is bounded, then by choosing C to be large enough in (2.7), we have $C\lambda b^{1/2}h(b)/(n - b/\mu_1) \geq 1$ and (2.7) is trivial. Therefore, in the following we can assume b is larger than any given constant. Moreover, since we assume $h(b) = O(b^{1/2})$ in the theorem, by choosing C to be large enough and c to be small enough in (2.7), we only need to consider the case where $h(b)/b^{1/2}$ is smaller than any given positive constant. In particular, we can assume

$$\frac{h(b)}{b^{1/2}} \leq \min \left\{ \frac{2}{\mu_1}, \frac{2(\theta'_1 - \theta_1) \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)}{\mu_1^2} \right\} \quad (3.29)$$

for some $\theta'_1 \in \Theta$ and $\theta_1 < \theta'_1 < 2\theta_1$. We first prove several lemmas that will be used in the proof of Theorem 2.3.

Lemma 3.2. *Let $\{X_1, \dots, X_n\}$ be independent, identically distributed random variables with distribution function F that can be imbedded in an exponential family, as in (2.1). Let $\mathbb{E}X_1 = \mu_0 < 0$. Let $S_0 = 0$ and $S_i = \sum_{k=1}^i X_k$ for $1 \leq i \leq n$. Suppose there exist $\theta_1 > 0$ such that $\Psi(\theta_1) = 0$. Let $\mathcal{F}_n = \sigma\{X_1, \dots, X_n\}$, and let T be a stopping time with respect to $\{\mathcal{F}_n\}$. Then we have*

$$\mathbb{P}(F \cap \{T < \infty\}) = \mathbb{E}_{\theta_1} [e^{-\theta_1 S_T} \mathbf{I}(F \cap \{T < \infty\})] \quad (3.30)$$

for any $F \in \mathcal{F}_T$.

Proof. Equation (3.30) follows by a direct application of Wald's likelihood ratio identity (cf. Theorem 1.1 of Woodroffe (1982)) to the sequence $\{X_1, X_2, \dots\}$. \square

Lemma 3.3. *Let $t = \lceil \frac{b}{\mu_1} + b^{1/2}h(b) \rceil$. We have*

$$\mathbb{P}_{\theta_1}(T_b > t) \leq Ce^{-ch^2(b)}.$$

Proof. Let

$$r = \frac{\mu_1^2}{2 \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)} h(b)/b^{1/2}.$$

By (3.29), we have $r < \theta_1$ and $\mu_1 r - \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta) r^2 / 2 \geq \mu_1 r / 2$. By

Markov's inequality and Taylor's expansion,

$$\begin{aligned}
\mathbb{P}_{\theta_1}(T_b > t) &\leq \mathbb{P}_{\theta_1}(S_t \leq b) \leq e^{rb} \mathbb{E}_{\theta_1} e^{-rS_t} \\
&\leq \exp \{rb - [\Psi(\theta_1) - \Psi(\theta_1 - r)]t\} \\
&\leq \exp \{rb - [\mu_1 r - \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)r^2/2]t\} \\
&\leq \exp \left\{ \frac{\sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)r^2b}{2\mu_1} - [\mu_1 r - \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)r^2/2]b^{1/2}h(b) \right\} \\
&\leq \exp \left\{ \frac{\sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)r^2b}{2\mu_1} - \frac{\mu_1 r}{2}b^{1/2}h(b) \right\} \\
&\leq \exp \left\{ - \frac{\mu_1^3}{8 \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)}h^2(b) \right\}.
\end{aligned}$$

This proves Lemma 3.3. \square

Lemma 3.4. *For positive integers m , we have*

$$\sum_{i=m}^{\infty} \mathbb{P}(S_i \geq 0) \leq Ce^{-cm}.$$

Proof. Lemma 3.4 follows from

$$\mathbb{P}(S_i \geq 0) \leq \mathbb{E}e^{\theta^* S_i} = e^{\Psi(\theta^*)i},$$

where $0 < \theta^* < \theta_1$ and $\Psi(\theta^*) < 0$. \square

Lemma 3.5. *Let $t_1 = \lfloor \frac{b}{\mu_1} - b^{1/2}h(b) \rfloor$. We have*

$$\mathbb{E}I(\cup_{0 \leq i < j \leq t_1} \{S_j - S_i \geq b\}) \leq Ce^{-\theta_1 b} \frac{b}{h^2(b)} e^{-ch^2(b)}.$$

Proof. We only need to consider the case when $t_1 > 0$. Let

$$r = \frac{\mu_1^2}{2 \sup_{\theta_1 \leq \theta \leq \theta'_1} \Psi''(\theta)} h(b)/b^{1/2}.$$

By (3.29), $\theta_1 + r \leq \theta'_1 \in \Theta$. We have

$$\mathbb{P}_{\theta_1}(S_j \geq b) \leq \exp \{j[\Psi(\theta_1 + r) - \Psi(\theta_1)] - rb\},$$

thus

$$\sum_{j=1}^i \mathbb{P}_{\theta_1}(S_j \geq b) \leq \frac{1}{1 - e^{\Psi(\theta_1) - \Psi(\theta_1 + r)}} \exp \{i[\Psi(\theta_1 + r) - \Psi(\theta_1)] - rb\}.$$

By (3.30) and Taylor's expansion,

$$\begin{aligned}
\mathbb{E}I(\cup_{0 \leq i < j \leq t_1} \{S_j - S_i \geq b\}) &\leq e^{-\theta_1 b} \sum_{i=1}^{t_1} \sum_{j=1}^i \mathbb{P}_{\theta_1}(S_j \geq b) \\
&\leq \frac{e^{-\theta_1 b}}{1 - e^{\Psi(\theta_1) - \Psi(\theta_1 + r)}} \sum_{i=1}^{t_1} \exp \{i[\Psi(\theta_1 + r) - \Psi(\theta_1)] - rb\} \\
&\leq e^{-\theta_1 b} \left(\frac{1}{1 - e^{\Psi(\theta_1) - \Psi(\theta_1 + r)}} \right)^2 \exp \{t_1[\Psi(\theta_1 + r) - \Psi(\theta_1)] - rb\} \\
&\leq Ce^{-\theta_1 b} \frac{b}{h^2(b)} \exp \{t_1[\Psi(\theta_1 + r) - \Psi(\theta_1)] - rb\} \\
&\leq Ce^{-\theta_1 b} \frac{b}{h^2(b)} \exp \left\{ \left(\frac{b}{\mu_1} - b^{1/2} h(b) \right) (r\mu_1 + \frac{\sup_{\theta_1 \leq \theta \leq \theta'_1} \Psi''(\theta)}{2} r^2) - rb \right\} \\
&\leq Ce^{-\theta_1 b} \frac{b}{h^2(b)} \exp \left\{ -b^{1/2} h(b) r\mu_1 + \frac{\sup_{\theta_1 \leq \theta \leq \theta'_1} \Psi''(\theta) b}{2\mu_1} r^2 \right\} \\
&\leq Ce^{-\theta_1 b} \frac{b}{h^2(b)} e^{-ch^2(b)}.
\end{aligned}$$

□

Lemma 3.6. *If $\int_{-\infty}^{\infty} |\varphi_{\theta_1}(t)| dt < \infty$ where $\varphi_{\theta_1}(t) = \mathbb{E}_{\theta_1} e^{itX_1}$, then S_{τ_+} under F_{θ_1} has bounded density and is strongly nonarithmetic in the sense that*

$$\liminf_{|\lambda| \rightarrow \infty} |1 - \varphi_{\theta_1}(\lambda)| > 0, \text{ where } \varphi_{\theta_1}(\lambda) = \mathbb{E}_{\theta_1} e^{i\lambda S_{\tau_+}}.$$

Proof. The condition $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$ implies that X_1 is strongly nonarithmetic. By (8.42) of Siegmund (1985) with $s = 1$, the distribution of S_{τ_+} is also strongly nonarithmetic. The condition $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$ also implies that the density of X_1 is bounded by a constant M . Therefore,

$$\begin{aligned}
\mathbb{P}_{\theta_1}(S_{\tau_+} \in [x, x + dx]) &\leq \mathbb{E}_{\theta_1} \sum_{n=0}^{\infty} I(S_1, \dots, S_n \leq 0, S_{n+1} \in [x, x + dx]) \\
&\leq \mathbb{E}_{\theta_1} \sum_{n=0}^{\infty} I(S_n \leq 0, S_{n+1} \in [x, x + dx]) \\
&= \sum_{n=0}^{\infty} \int_{-\infty}^0 \mathbb{P}_{\theta_1}(S_n = dt) \mathbb{P}_{\theta_1}(X_1 \in [x - t, x + dx - t]) \\
&\leq M dx \sum_{n=0}^{\infty} \mathbb{P}_{\theta_1}(S_n \leq 0) \leq C dx,
\end{aligned}$$

where in the last inequality we used

$$\mathbb{P}_{\theta_1}(S_n \leq 0) \leq e^{\Psi(\theta_1 - \theta^*)n} \quad (3.31)$$

for $0 < \theta^* < \theta_1$ so that $\Psi(\theta_1 - \theta^*) < 0$. This proves that S_{τ_+} under F_{θ_1} has bounded density. \square

Proof of Theorem 2.3. We embed the sequence $\{X_1, \dots, X_n\}$ into an infinite i.i.d. sequence $\{\dots, X_{-1}, X_0, X_1, \dots\}$. For a positive integer m , let ω_m^+ be the m -shifted sample path of $\omega := \{X_1, \dots, X_n\}$, so $S_i(\omega_m^+) = S_{m+i}(\omega) - S_m(\omega)$, $T_b(\omega_m^+) = \inf\{n \geq 1 : S_n(\omega_m^+) \notin [0, b]\}$, and $\tau_b(\omega_m^+)$, $\tau_+(\omega_m^+)$ are defined similarly. Let $t = \lceil \frac{b}{\mu_1} + b^{1/2}h(b) \rceil$ and $m < t$ to be chosen at the end of this proof. For $1 \leq \alpha \leq n - t$, let

$$Y_\alpha = \mathbf{I}(S_\alpha < S_{\alpha-\beta}, \forall 1 \leq \beta \leq m; T_b(\omega_\alpha^+) \leq t, S_{T_b}(\omega_\alpha^+) \geq b).$$

That is, Y_α is the indicator of the event that the sequence $\{S_i\}$ reaches a local minimum at α and the α -shifted sequence $\{S_i(\omega_\alpha^+)\}$ exits the interval $[0, b]$ within time t and the first exiting position is b . Let

$$W = \sum_{\alpha=1}^{n-t} Y_\alpha.$$

In the following, we first compare $p_{n,b}$ with $\mathbb{P}(W \geq 1)$. Then, we approximate the distribution of W by the Poisson distribution with mean $\mathbb{E}(W)$. Finally, we calculate approximately $\mathbb{E}(W)$.

First, from the definition of W , we have $p_{n,b} \geq \mathbb{P}(W \geq 1)$ and with $t_1 = \lfloor b/\mu_1 - b^{1/2}h(b) \rfloor$,

$$\begin{aligned} & \{ \max_{0 \leq i < j \leq n} (S_j - S_i) \geq b \} \setminus \{W \geq 1\} \\ & \subset (\cup_{k=0}^{n-t-1} \{S_{T_b}(\omega_k^+) \geq b, T_b(\omega_k^+) > t\}) \\ & \quad \cup (\cup_{k \in [0, m] \cup (n-t, n-t_1)} \{S_{T_b}(\omega_k^+) \geq b, T_b(\omega_k^+) \leq t\}) \\ & \quad \cup (\cup_{n-t_1 \leq i < j \leq n} \{S_j - S_i \geq b\}). \end{aligned}$$

By symmetry,

$$\begin{aligned} & p_{n,b} - \mathbb{P}(W \geq 1) \\ & \leq (n-t)\mathbb{P}(S_{T_b} \geq b, T_b > t) + (m + 2b^{1/2}h(b) + 2)\mathbb{P}(S_{T_b} \geq b, T_b \leq t) \\ & \quad + \mathbb{E}\mathbf{I}(\cup_{0 \leq i < j \leq t_1} \{S_j - S_i \geq b\}). \end{aligned} \tag{3.32}$$

By (3.30) and Lemma 3.3, we have

$$\mathbb{P}(S_{T_b} \geq b) = \mathbb{E}_{\theta_1}[e^{-\theta_1 S_{T_b}} \mathbf{I}(S_{T_b} \geq b)] \leq e^{-\theta_1 b} \tag{3.33}$$

and

$$\begin{aligned} & \mathbb{P}(S_{T_b} \geq b, T_b > t) = \mathbb{E}_{\theta_1}[e^{-\theta_1 S_{T_b}} \mathbf{I}(S_{T_b} \geq b, T_b > t)] \\ & \leq e^{-\theta_1 b} \mathbb{P}_{\theta_1}(T_b > t) \leq C e^{-\theta_1 b - ch^2(b)}. \end{aligned} \tag{3.34}$$

Along with Lemma 3.5,

$$\begin{aligned} p_{n,b} - \mathbb{P}(W \geq 1) \\ \leq C(n - b/\mu_1)e^{-\theta_1 b} \left\{ e^{-ch^2(b)} + \frac{m + b^{1/2}h(b)}{n - b/\mu_1} + \frac{b/h^2(b)}{n - b/\mu_1} e^{-ch^2(b)} \right\}. \end{aligned} \quad (3.35)$$

Next, we use Theorem 3.1 to obtain a bound on the total variation distance between the distribution of W and $Poi(\lambda_1)$ with $\lambda_1 := \mathbb{E}(W) = (n - t)\mathbb{E}Y_\alpha$. For each $1 \leq \alpha \leq n - t$, let $B_\alpha = \{1 \leq \beta \leq n - t : |\beta - \alpha| \leq t + m\}$. In applying Theorem 3.1, by our definition of B_α , $b_3 = 0$. Since $|B_\alpha| \leq 2(t + m) + 1$, we have

$$b_1 < [2(t + m) + 1]\lambda_1 \mathbb{E}Y_\alpha \leq C(n - t)(t + m)\mathbb{P}^2(S_{T_b} \geq b) \leq C(n - t)(t + m)e^{-2\theta_1 b}. \quad (3.36)$$

Let

$$\tilde{Y}_\alpha = \mathbb{I}(T_b(\omega_\alpha^+) \leq t, S_{T_b}(\omega_\alpha^+) \geq b).$$

We have for b_2 in (3.2),

$$\begin{aligned} b_2 &\leq \sum_{\alpha=1}^{n-t} \sum_{\alpha \neq \beta \in B_\alpha} \mathbb{E}(Y_\alpha Y_\beta) \\ &\leq 2 \sum_{\beta=1}^{n-t} \left[\sum_{\beta-t-m \leq \alpha < \beta-m} \mathbb{E}Y_\beta \tilde{Y}_\alpha + \sum_{\beta-m \leq \alpha \leq \beta-1} \mathbb{E}Y_\beta \tilde{Y}_\alpha \right]. \end{aligned}$$

For $\beta - t - m \leq \alpha < \beta - m$, because $S_\beta < S_\alpha$ implies $T_b(\omega_\alpha^+) \leq \beta - \alpha$, we have

$$\begin{aligned} \mathbb{E}Y_\beta \tilde{Y}_\alpha &= \mathbb{E}Y_\beta \tilde{Y}_\alpha [\mathbb{I}(S_\beta \geq S_\alpha) + \mathbb{I}(S_\beta < S_\alpha)] \\ &\leq \mathbb{E}[\mathbb{I}(S_\beta - S_\alpha \geq 0) \tilde{Y}_\beta + \mathbb{E}[\mathbb{I}(S_{T_b}(\omega_\alpha^+) \geq b, T_b(\omega_\alpha^+) \leq \beta - \alpha) \tilde{Y}_\beta]. \end{aligned}$$

By independence and symmetry,

$$\sum_{\beta-t-m \leq \alpha < \beta-m} \mathbb{E}Y_\beta \tilde{Y}_\alpha \leq \mathbb{E}\tilde{Y}_1 \sum_{i=m}^{t+m} [\mathbb{P}(S_i \geq 0) + \mathbb{P}(S_{T_b} \geq b, T_b \leq t + m)].$$

For $\beta - m \leq \alpha \leq \beta - 1$, because $Y_\beta = 1$ implies $S_\alpha > S_\beta$, which in turn implies $T_b(\omega_\alpha^+) \leq \beta - \alpha$, we have

$$\begin{aligned} \sum_{\beta-m \leq \alpha \leq \beta-1} \mathbb{E}Y_\beta \tilde{Y}_\alpha &\leq \sum_{\beta-m \leq \alpha \leq \beta-1} \mathbb{E}\tilde{Y}_\beta \mathbb{I}(S_{T_b}(\omega_\alpha^+) \geq b, T_b(\omega_\alpha^+) \leq \beta - \alpha) \\ &\leq \mathbb{E}\tilde{Y}_1 \sum_{i=1}^m \mathbb{P}(S_{T_b} \geq b, T_b \leq i). \end{aligned}$$

Therefore,

$$\begin{aligned}
b_2 &\leq 2(n-t)\mathbb{E}\tilde{Y}_1 \left[\sum_{i=m}^{t+m} (\mathbb{P}(S_i \geq 0) + \mathbb{P}(S_{T_b} \geq b, T_b \leq t+m)) \right. \\
&\quad \left. + \sum_{i=1}^m \mathbb{P}(S_{T_b} \geq b, T_b \leq i) \right] \\
&\leq 2(n-t)e^{-\theta_1 b} [Ce^{-cm} + (t+m)e^{-\theta_1 b}]
\end{aligned} \tag{3.37}$$

by Lemma 3.4 and (3.33). From (3.1), (3.36) and (3.37),

$$|\mathbb{P}(W \geq 1) - (1 - e^{-\lambda_1})| \leq C(n-t)e^{-\theta_1 b} [(t+m)e^{-\theta_1 b} + e^{-cm}]. \tag{3.38}$$

Now we bound the difference between λ_1 and

$$\lambda_2 = (n-t)\mathbb{P}(\tau_0 = \infty)\mathbb{P}(S_{T_b} \geq b).$$

Recall

$$\begin{aligned}
\lambda_1 &= (n-t)\mathbb{E}Y_\alpha \\
&= (n-t)\mathbb{P}(S_{\alpha-\beta} - S_\alpha < 0, \forall 1 \leq \beta \leq m)\mathbb{P}(T_b(\omega_\alpha^+) \leq t, S_{T_b}(\omega_\alpha^+) \geq b) \\
&= (n-t)\mathbb{P}(\tau_0 > m)\mathbb{P}(T_b \leq t, S_{T_b} \geq b).
\end{aligned}$$

From the upper and lower bounds of their difference

$$\begin{aligned}
\lambda_2 - \lambda_1 &\leq (n-t)\mathbb{P}(T_b > t, S_{T_b} \geq b), \\
\lambda_1 - \lambda_2 &\leq (n-t)\mathbb{P}(S_{T_b} \geq b)\mathbb{P}(m < \tau_0 < \infty),
\end{aligned}$$

we have

$$\begin{aligned}
|\lambda_1 - \lambda_2| &\leq C(n-t)e^{-\theta_1 b - ch^2(b)} + (n-t)e^{-\theta_1 b} \sum_{i=m}^{\infty} \mathbb{P}(S_i \geq 0) \\
&\leq C(n-t)e^{-\theta_1 b} [e^{-ch^2(b)} + e^{-cm}]
\end{aligned} \tag{3.39}$$

by (3.34), (3.33) and Lemma 3.3.

Finally we calculate approximately λ_2 . By (3.30),

$$\lambda_2 = (n-t)e^{-\theta_1 b}\mathbb{P}(\tau_0 = \infty)\mathbb{E}_{\theta_1}(e^{-\theta_1(S_{T_b}-b)}, S_{T_b} \geq b). \tag{3.40}$$

Since

$$\begin{aligned}
\mathbb{E}_{\theta_1}(e^{-\theta_1(S_{T_b}-b)}) &= \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{T_b}-b)}, S_{T_b} \geq b) \\
&\quad + \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{T_b}-b)}, S_{T_b} < 0),
\end{aligned}$$

we have

$$\begin{aligned}
& \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{T_b}-b)}, S_{T_b} \geq b) \\
&= \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b}-b)}) - \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b}-b)}, S_{T_b} < 0) \\
&= \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b}-b)}) - \mathbb{E}_{\theta_1}\left\{\mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b}-b)}|S_{T_b}), S_{T_b} < 0\right\}.
\end{aligned} \tag{3.41}$$

We first consider the non-arithmetic case. Let $\tau_+^{(0)} = 0$, and let $\tau_+^{(k)}$ be defined recursively as $\tau_+^{(k+1)} = \inf\{n > \tau_+^{(k)} : S_n > S_{\tau_+^{(k)}}\}$. Define $U(x) = \sum_{k=0}^{\infty} \mathbb{P}_{\theta_1}(S_{\tau_+^{(k)}} \leq x)$. Observe that $\{S_{\tau_+^{(k+1)}} - S_{\tau_+^{(k)}}, k = 0, 1, \dots\}$ are i.i.d. with the same distribution as S_{τ_+} . By Lemma 3.6 and (2) of Stone (1965),

$$U(x) = \frac{x}{\mathbb{E}_{\theta_1} S_{\tau_+}} + \frac{\mathbb{E}_{\theta_1}(S_{\tau_+}^2)}{\mathbb{E}_{\theta_1} S_{\tau_+}} + o(e^{-cx}), \text{ as } x \rightarrow \infty. \tag{3.42}$$

Following the proof of Corollary 8.33 of Siegmund (1985), we have for $x \geq 0$,

$$\begin{aligned}
& \mathbb{P}_{\theta_1}(S_{\tau_b} - b > x) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_{\theta_1}(S_{\tau_+^{(n)}} < b, S_{\tau_+^{(n+1)}} > b + x) \\
&= \left(\int_{(0, b/2]} + \int_{(b/2, b)}\right) U(dt) \mathbb{P}_{\theta_1}(S_{\tau_+} > b + x - t) \\
&= O(\mathbb{P}_{\theta_1}(S_{\tau_+} > b/2)U(b/2)) + \int_{(b/2, b)} U(dt) \mathbb{P}_{\theta_1}(S_{\tau_+} > b + x - t).
\end{aligned} \tag{3.43}$$

For $x > 0$,

$$\begin{aligned}
\mathbb{P}_{\theta_1}(S_{\tau_+} > x) &= \mathbb{E}_{\theta_1}\left[\sum_{i=0}^{\infty} \mathbf{I}(S_0, \dots, S_i \leq 0, X_{i+1} > x - S_i)\right] \\
&\leq \sum_{i=0}^{\infty} \mathbb{P}_{\theta_1}(S_i \leq 0, X_{i+1} > x) \\
&= \sum_{i=0}^{\infty} \mathbb{P}_{\theta_1}(S_i \leq 0) \mathbb{P}_{\theta_1}(X_{i+1} > x) \leq C \mathbb{P}_{\theta_1}(X_1 > x)
\end{aligned}$$

where we used (3.31). Therefore, the right tail probability of S_{τ_+} under F_{θ_1} decays exponentially. Along with (3.42), the first term on the right-hand side of (3.43) is bounded by $o(e^{-cb})$. Let $j = \lceil e^{cb} \rceil$ with small enough c , and let $\Delta = \frac{b}{2j}$. Then

$$\int_{(b/2, b)} U(dt) \mathbb{P}_{\theta_1}(S_{\tau_+} > b + x - t) \geq A$$

where

$$A = \sum_{k=1}^j [U(b - (k-1)\Delta) - U(b - k\Delta)] \mathbb{P}_{\theta_1}(S_{\tau_+} > x + k\Delta),$$

and by (3.42) and the fact that S_{τ_+} under F_{θ_1} has bounded density (cf. Lemma 3.6),

$$\begin{aligned} & \int_{(b/2, b)} U(dt) \mathbb{P}_{\theta_1}(S_{\tau_+} > b + x - t) - A \\ & \leq \sum_{k=1}^j [U(b - (k-1)\Delta) - U(b - k\Delta)] \mathbb{P}_{\theta_1}(S_{\tau_+} \in [x + (k-1)\Delta, x + k\Delta]) \\ & = o(e^{-cb}). \end{aligned}$$

From (3.42),

$$A = \sum_{k=1}^j \frac{\Delta}{\mathbb{E}_{\theta_1} S_{\tau_+}} \mathbb{P}_{\theta_1}(S_{\tau_+} > x + k\Delta) + O(je^{-cb})$$

with the same c as in (3.42). By choosing c in the definition of j to be small enough, we have $je^{-cb} = o(e^{-cb})$. Using the fact that S_{τ_+} under F_{θ_1} has bounded density and an exponential tail, we have

$$\sum_{k=1}^j \frac{\Delta}{\mathbb{E}_{\theta_1} S_{\tau_+}} \mathbb{P}_{\theta_1}(S_{\tau_+} > x + k\Delta) = \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \int_x^\infty \mathbb{P}_{\theta_1}(S_{\tau_+} > y) dy + o(e^{-cb}).$$

Therefore,

$$\begin{aligned} A &= \sum_{k=1}^j \left[\frac{\Delta}{\mathbb{E}_{\theta_1} S_{\tau_+}} + o(e^{-cb}) \right] \mathbb{P}_{\theta_1}(S_{\tau_+} > x + k\Delta) \\ &= \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \int_x^\infty \mathbb{P}_{\theta_1}(S_{\tau_+} > y) dy + o(e^{-cb}). \end{aligned}$$

By (3.43) and the above argument,

$$\mathbb{P}_{\theta_1}(S_{\tau_b} - b > x) = \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \int_x^\infty \mathbb{P}_{\theta_1}(S_{\tau_+} > y) dy + o(e^{-cb}). \quad (3.44)$$

Using the integration by parts formula and the above equality,

$$\begin{aligned}
& \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b}-b)}) \\
&= 1 - \theta_1 \int_0^\infty \mathbb{P}_{\theta_1}(S_{\tau_b} - b > x) e^{-\theta_1 x} dx \\
&= 1 - \frac{\theta_1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \int_0^\infty \int_x^\infty \mathbb{P}_{\theta_1}(S_{\tau_+} > y) e^{-\theta_1 x} dy dx + o(e^{-cb}) \\
&= 1 + \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \int_0^\infty (e^{-\theta_1 y} - 1) \mathbb{P}_{\theta_1}(S_{\tau_+} > y) dy + o(e^{-cb}) \tag{3.45} \\
&= \frac{1}{\mu_1 \mathbb{E}_{\theta_1} \tau_+} \int_0^\infty e^{-\theta_1 y} \mathbb{P}_{\theta_1}(S_{\tau_+} > y) dy + o(e^{-cb}) \\
&= \frac{1}{\theta_1 \mu_1} \exp \left(- \sum_{k=1}^\infty \frac{1}{k} \mathbb{E}_{\theta_1} e^{-\theta_1 S_k^+} \right) + o(e^{-cb}),
\end{aligned}$$

where in the last equality we used (3.25). From (3.41) and (3.45), we have, with $\tau_- := \inf\{n : S_n < 0\}$,

$$\begin{aligned}
& \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{T_b}-b)}, S_{T_b} \geq b) \\
&= \frac{1}{\theta_1 \mu_1} \exp \left(- \sum_{n=1}^\infty \frac{1}{n} \mathbb{E}_{\theta_1} e^{-\theta_1 S_n^+} \right) \mathbb{P}_{\theta_1}(S_{T_b} \geq b) + o(e^{-cb}) \\
&= \frac{1}{\theta_1 \mu_1} \exp \left(- \sum_{n=1}^\infty \frac{1}{n} \mathbb{E}_{\theta_1} e^{-\theta_1 S_n^+} \right) \mathbb{P}_{\theta_1}(\tau_- = \infty) + o(e^{-cb})
\end{aligned}$$

as $b \rightarrow \infty$, where we used

$$\begin{aligned}
0 &\leq \mathbb{P}_{\theta_1}(S_{T_b} \geq b) - \mathbb{P}_{\theta_1}(\tau_- = \infty) \\
&\leq \sum_{i=1}^\infty \mathbb{P}_{\theta_1}(S_i < -b) \leq e^{-\theta^* b} \sum_{i=1}^\infty e^{\Psi(\theta_1 - \theta^*)i} = o(e^{-cb}),
\end{aligned}$$

where $0 < \theta^* < \theta_1$ so that $\Psi(\theta_1 - \theta^*) < 0$.

By (3.40) and

$$\begin{aligned}
\mathbb{P}(\tau_0 = \infty) \mathbb{P}_{\theta_1}(\tau_- = \infty) &= \exp \left\{ - \sum_{k=1}^\infty \frac{1}{k} [\mathbb{P}(S_k \geq 0) + \mathbb{P}_{\theta_1}(S_k < 0)] \right\} \\
&= \exp \left\{ - \sum_{k=1}^\infty \frac{1}{k} [\mathbb{P}_{\theta_1}(e^{-\theta_1 S_k}, S_k \geq 0) + \mathbb{P}_{\theta_1}(S_k < 0)] \right\} \\
&= \exp \left\{ - \sum_{k=1}^\infty \frac{1}{k} \mathbb{E}_{\theta_1} e^{-\theta_1 S_k^+} \right\}
\end{aligned}$$

from Lemma 3.2 and Corollary 2.4 of Woodroffe (1982), we have,

$$\begin{aligned}\lambda_2 &= (n-t)e^{-\theta_1 b} \left\{ \frac{1}{\theta_1 \mu_1} \exp \left(-2 \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}_{\theta_1} e^{-\theta_1 S_n^+} \right) + o(e^{-cb}) \right\} \\ &= \left[1 + O\left(\frac{b^{1/2} h(b)}{n - b/\mu_1} \right) \right] \lambda + (n-t)e^{-\theta_1 b} o(e^{-cb}).\end{aligned}\quad (3.46)$$

For the arithmetic case, assume X_1 is integer valued with span 1, and b is an integer. By a similar and simpler argument as for (3.44), we have, for integers $k \geq 0$,

$$\begin{aligned}\mathbb{P}_{\theta_1}(S_{\tau_b} - b = k) &= \sum_{n=0}^{\infty} \mathbb{P}_{\theta_1}(S_{\tau_+^{(n)}} < b, S_{\tau_+^{(n+1)}} = b+k) \\ &= \sum_{m=1}^{b-1} \left[\sum_{n=0}^{\infty} \mathbb{P}_{\theta_1}(S_{\tau_+^{(n)}} = m) \right] \mathbb{P}_{\theta_1}(S_{\tau_+} = b+k-m) \\ &= O\left(\sum_{m=1}^{\lfloor b/2 \rfloor} \sum_{n=0}^{\infty} \mathbb{P}_{\theta_1}(S_{\tau_+^{(n)}} = m) \mathbb{P}_{\theta_1}(S_{\tau_+} \geq \lfloor b/2 \rfloor) \right) \\ &\quad + \sum_{m=\lfloor b/2 \rfloor + 1}^{b-1} \mathbb{P}_{\theta_1}(S_{\tau_+} = b+k-m) \left(\frac{1}{\mathbb{E}_{\theta_1}(S_{\tau_+})} + o(e^{-cb}) \right) \\ &= \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \mathbb{P}_{\theta_1}(S_{\tau_+} > k) + o(e^{-cb}).\end{aligned}$$

By the above equality and (3.28),

$$\begin{aligned}\mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b} - b)}) &= \sum_{k=0}^{\infty} e^{-\theta_1 k} \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \mathbb{P}_{\theta_1}(S_{\tau_+} > k) + o(e^{-cb}) \\ &= \frac{1}{\mu_1 \mathbb{E}_{\theta_1} \tau_+ (1 - e^{-\theta_1})} [1 - \mathbb{E}_{\theta_1} e^{-\theta_1 S_{\tau_+}}] + o(e^{-cb}) \\ &= \frac{1}{\mu_1 (1 - e^{-\theta_1})} \exp \left(- \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}_{\theta_1} e^{-\theta_1 S_n^+} \right) + o(e^{-cb}).\end{aligned}$$

Similar calculation as for the non-arithmetic case yield

$$\lambda_2 = (n-t)e^{-\theta_1 b} \left\{ \frac{1}{(1 - e^{-\theta_1}) \mu_1} \exp \left(-2 \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}_{\theta_1} e^{-\theta_1 S_n^+} \right) + o(e^{-cb}) \right\}. \quad (3.47)$$

Theorem 2.3 is proved by combining (3.35), (3.38), (3.39), (3.46) and (3.47) and letting $m = \lfloor ch^2(b) \rfloor$ such that $m < t$. \square

4 DISCUSSION

The arguments we used to prove Theorem 2.1 and Theorem 2.3 may be useful in proving rates of convergence for tail probabilities of other test statistics for detecting local signals in sequences of independent random variables. Two for which some new techniques will be needed are the Levin and Kline statistic (Levin and Kline (1985)) and the generalized likelihood ratio statistic.

For example, let $\{X_1, \dots, X_n\}$ be independent random variables from the exponential family (2.1). Consider the testing problem at the beginning of the introduction. If the mean of X_1 is known and without loss of generality equal to 0, the generalized likelihood ratio statistic is $\max_{1 \leq i < j < n} \sup_{\theta} [\theta(S_j - S_i) - (j - i)\Psi(\theta)]$, where we have assumed without loss of generality that $\Psi(0) = 0 = \dot{\Psi}(0)$. Siegmund and Venkatraman (1995) derived an asymptotic approximation for the tail probability of this statistic in the normal case, while Siegmund and Yakir (2000) obtained similar results for a general exponential family.

If the mean of X_1 is unknown, the statistic is more complicated; and its tail probability should be evaluated conditionally, given the value of S_n , which is a sufficient statistic for the unknown value of θ under the null hypothesis of no change-point.

Enough is known about the first order asymptotic behavior of these statistics to permit hypotheses about the rate of convergence; but a proof will involve new techniques.

ACKNOWLEDGMENTS

XF was supported by NUS-Overseas Postdoctoral Fellowship from the National University of Singapore. DS was partially supported by the National Science Foundation.

REFERENCES

- R. Arratia, L. Goldstein and L. Gordon (1990). Poisson approximation and the Chen-Stein method. With comments and a rejoinder by the authors. *Statist. Sci.* **5**, 403–434.
- R. Arratia, L. Gordon and M. S. Waterman (1990). The Erdős-Rényi law in distribution, for coin tossing and sequence matching. *Ann. Statist.* **18**, 539–570.
- A. D. Barbour and L. H. Y. Chen (2005). *An Introduction to Stein's method*. Lecture Notes Series **4**, Institute for Mathematical Sciences, National University of Singapore, Singapore University Press and World Scientific.
- H. P. Chan and N. R. Zhang (2007). Scan statistics with weighted observations. *J. Amer. Statist. Assoc.* **102**, 595–602.
- L. H. Y. Chen (1975). An approximation theorem for sums of certain randomly selected indicators. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **33**, 69–74.

- P. Diaconis and D. A. Freedman (1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. *J. Theoret. Probab.* **1**, 381–410.
- W. Feller (1971). *Probability Theory and Its Applications*, Vol. II, 2nd ed., Wiley, New York.
- J. Glaz, J. Naus and S. Wallenstein (2001). *Scan statistics*, Springer-Verlag, New York.
- D. Iglehart (1971). Functional limit theorems for the queue $GI/G/1$ in light traffic. *Advances in Appl. Probability* **3**, 269–281.
- S. Karlin and Y. Rinott (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **4**, 467–498.
- J. Komlós and G. Tusnády (1975). On sequences of “pure heads”. *Ann. Probab.* **3**, 608–617.
- B. Levin and J. Kline (1985). The cusum test of homogeneity with an application in spontaneous abortion epidemiology. *Statist. Med.* **4**, 469–488.
- V. V. Petrov (1965). On the probabilities of large deviations for sums of independent random variables. (Russian) *Teor. Veroyatnost. i Primenen* **10**, 310–322.
- D. Siegmund (1985). *Sequential analysis: Tests and confidence intervals*. Springer, New York.
- D. Siegmund (1988). Approximate tail probabilities for the maxima of some random fields. *Ann. Probab.* **16**, 487–501.
- D. Siegmund and E. S. Venkatraman (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.* **23**, 255–271.
- D. Siegmund and B. Yakir (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6**, 191–213.
- C. Stein (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Stat. Prob.* **2** Univ. California Press, Berkeley, Calif., 583–602.
- C. Stein (1986). *Approximate Computation of Expectations*. Lecture Notes 7, Inst. Math. Statist., Hayward, Calif.
- C. Stone (1965). On moment generating functions and renewal theory. *Ann. Math. Statist.* **36**, 1298–1301.
- M. Woodroffe (1982). *Nonlinear renewal theory in sequential analysis*. Society for Industrial and Applied Mathematics, Philadelphia.
- Q. Yao (1993). Tests for change-points with epidemic alternatives. *Biometrika* **80**, 179–191.